Erin Zaroukian & Jonathan Z. Bakdash
Human Research and Engineering Directorate
US Army Research Laboratory
Aberdeen Proving Ground, USA
erin.g.zaroukian.ctr@mail.mil jonathan.z.bakdash.civ@mail.mil

Alun Preece & Will Webberley
Crime & Security Research Institute
Cardiff University
Cardiff, UK
PreeceAD@cardiff.ac.uk WebberleyWM@cardiff.ac.uk
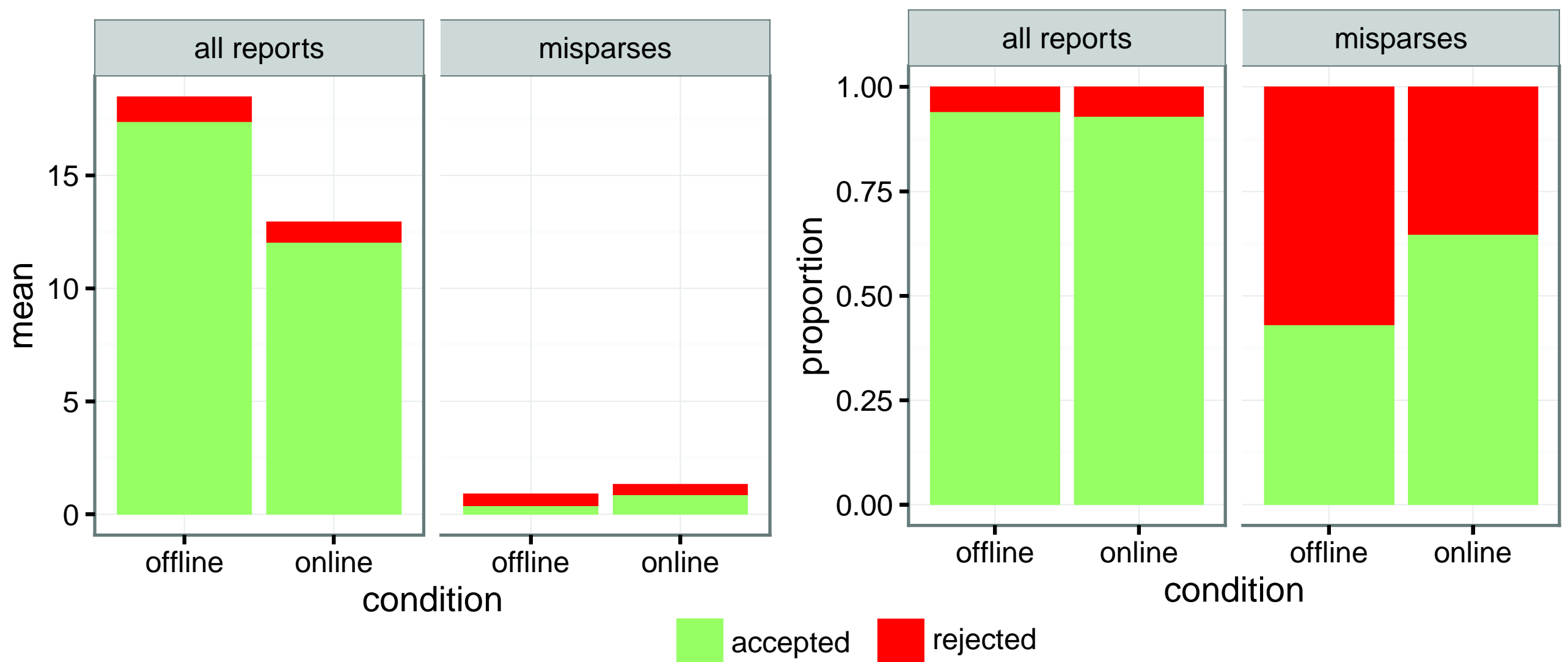
Abstract

- Users participated in a simulated intelligence, surveillance, and reconnaissance task using a conversational agent
- Automation bias and loss of SA? [1]-[2] Users accepted a surprising number of misparses of their inputs by the agent
- Highlights concerns for implementing a flexible conversational interface in safety critical environments (e.g., military, emergency operations)

Behavioral Experiment

- Used *Simple Human Experiments Regarding Locally Observed Collective Knowledge* (SHERLOCK) platform [3]-[5]

  Designed to support simple situation awareness tasks and the automated fusion of information from human tactical intelligent team members

  - 161 participants
  - 36 questions
  - Questions answered by visiting physical locations
  - Users report answers to conversational agent to populate knowledge base
  - Knowledge base reflected in dashboard

Offline condition (n=72)
Local knowledge base
(no common operating picture)

Online condition (n=89)
Shared knowledge base
(common operating picture)

Misparses

≡ parses containing a word/phrase that the computer agent was not preprogrammed with (computer agent was preprogrammed with every entity and relation needed to complete the task)

- Often – misspelling, use of an entity or relation that was not preprogrammed

User: Zebra is in the **solver** room →
Agent: there is a room named '**solver**' [should be "silver"]

  User confirmation creates new entity in the knowledge base, allowing for further misparses

User: The apple is in the **solver** room →
Agent: the fruit 'apple' is in the location '**solver**'
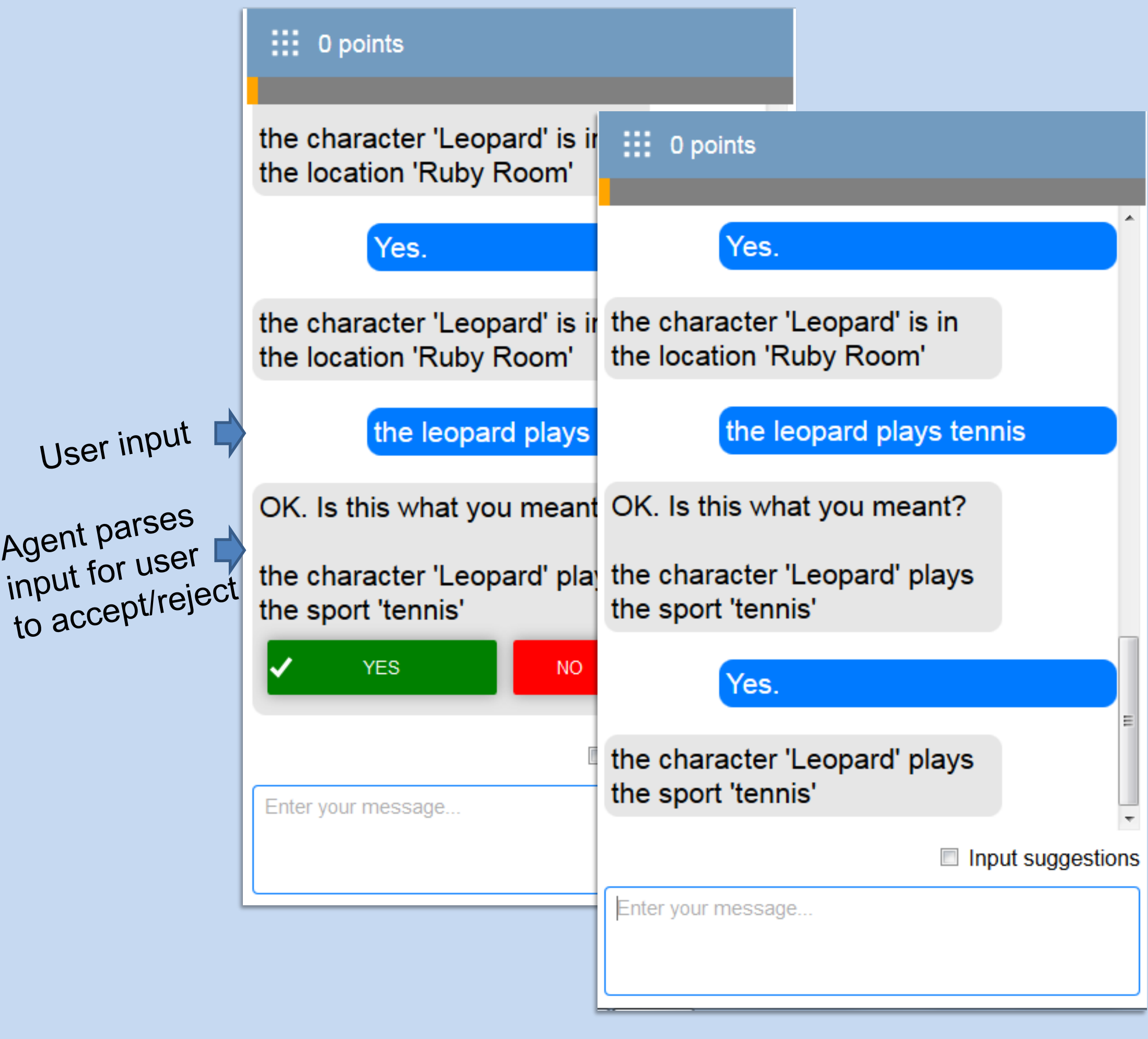
Conversational Interface

Results

- All reports
  - Offline groups submitted more reports overall
  - Accept rate similar across conditions
    (p = 0.307, Odds Ratio (OR) = 0.835, OR 95% CI = 0.605 — 1.152)
- Misparses
  - Online participants had a higher accept rate
    (p = 0.005, Odds Ratio (OR) = 2.407, OR 95% CI = 1.299 — 4.514)
  → The odds of a misparse being rejected in the offline condition are more than twice the odds in the online condition

Conclusions

Two potential complementary explanations for misparses:

1. Participants thought these parses were correct because of lack of familiarity with the conversational interface
   - Change the interface or training
   - Spell-check, Term merging, reserve words [6]
2. Automation bias and complacency: Users tended to accept the agent interpretation because it was generally correct (92%)
   - Hide common operating picture

[1] L. Bainbridge, "Ironies of automation," *Automatica*, vol. 19, no. 6, pp. 775–779, 1983.
[2] R. Parasuraman and D. H. Manzey, "Complacency and Bias in Human Use of Automation: An Attentional Integration," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 52, no. 3, pp. 381–410, Oct. 2010.
[3] A. Preece, W. Webberley, D. Braines, E. G. Zaroukian, and J. Z. Bakdash, "SHERLOCK: Experimental Evaluation of a Conversational Agent for Mobile Information Tasks," under review.
[4] A. Preece *et al.*, "SHERLOCK: Simple Human Experiments Regarding Locally Observed Collective Knowledge," U.S. Army Research Laboratory, Aberdeen Proving Ground, MD, Tech Rep. ARL-RP-0560, Dec 2015.
[5] A. D. Preece, W. Webberley, D. Braines, E. Zaroukian, and J. Bakdash, "Human computer collaboration at the edge: Enhancing collective situation understanding with controlled natural language," 2016.
[6] D. Weir, H. Pohl, S. Rogers, K. Vertanen, and P. O. Kristensson, "Uncertain text entry on mobile devices," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2014, pp. 2307– 2316.